Simplicial Homology and Topological Data Analysis

Patrick O'Neil

George Mason University

October 15, 2013

Overview

Data Analysis

Point Clouds

What is a Simplex?

A simplex is a generalization of a triangle to arbitrary dimensions.

Definition

A *k-simplex* is a k-dimensional polytope which is the convex hull of $k+1$ vertices.

For lower dimensions, these have familiar names: vertex (0), edge (1), triangle (2), tetrahedron (3).

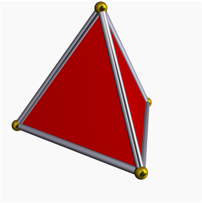


Figure: 3-simplex

Simplex: Abstract Representation

Given $k + 1$ vertices $\{v_0, v_1, ..., v_k\}$, we can denote the k-simplex formed by these vertices as $\sigma = (v_0, v_1, ..., v_k)$. A *face* of σ is any subset of σ.

Example: the 2-simplex (v_0, v_1, v_3) is a face of the 4-simplex $(v_0, v_1, v_2, v_3, v_4)$

Note that given an abstract k-simplex σ, it is possible to form a geometric realization of the complex. The result will be the convex hull of the vertices in σ, embedded in $\mathbb{R}^d$, $d \geq k$.

What is a Simplicial Complex?

One can form a *Simplicial Complex* by joining simplicies in a certain way.

Definition

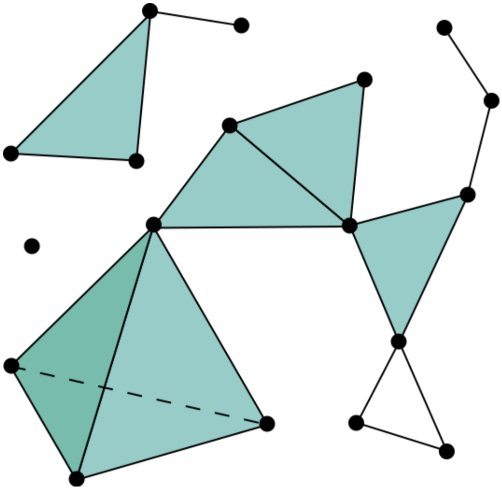
A *Simplicial Complex* $\mathcal{K}$ is a set of simplicies such that

a) $\mathcal{K}$ is closed under the subset relation: If $\tau \subseteq \sigma \in \mathcal{K}$, then $\tau \in \mathcal{K}$.

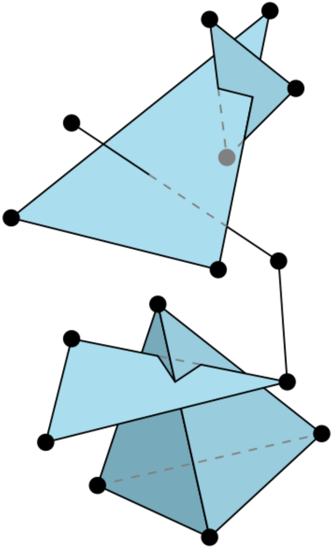
b) $(*)$ If $\sigma_1, \sigma_2 \in \mathcal{K}$, then $\sigma_1 \cap \sigma_2$ is a face of both σ_1 and σ_2.

We only need $(*)$ when discussing a geometric realization of a simplicial complex.

Simplicial Complex Example



Simplicial Complex Non-Example



Cover and Nerve: Definition

We now define an open cover and the associated nerve of the open cover:

Definition

Given a topological space $\mathbb{X}$, an *open cover* $\mathcal{O}$ of a set $S \subseteq \mathbb{X}$ is

$$\mathcal{O} = \{U_i\}_{i \in I}$$

where $U_i \subseteq \mathbb{X}$ and I is an indexing set, such that $S \subseteq \bigcup_{i \in I} U_i$

Definition

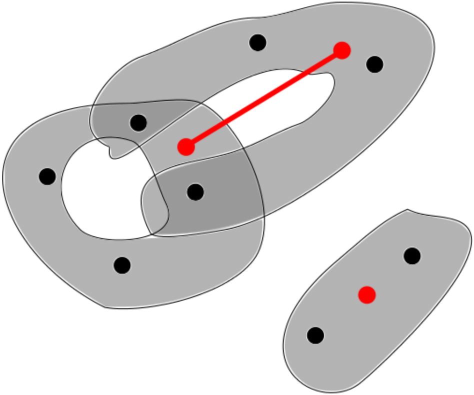
Given an open cover $\mathcal{O} = \{U_i\}_{i \in I}$, the *nerve* N of $\mathcal{O}$ is a subset of $\mathcal{P}(I)$ such that

a) $\emptyset \in N$

b) If $\cap_{j \in J} U_j \neq \emptyset$ for $J \subseteq I$, then $J \in N$

From these definitions, it is clear that the nerve of an open cover is a simplicial complex!

Cover and Nerve: Example



Pretty simple simplicial complex.

Open Cover Dependence

The simplicial complex given by the nerve is completely dependent upon the open cover that is chosen.

There are **many** ways to choose an open cover. We will present the *Voronoi Diagram* which produces the *Delaunay Complex*.

We will then present the Alpha complex which allows us to analyze the data at different levels of detail.

Delaunay Complex: Definition

Given a subset S of a metric space $\mathbb{X} = (\mathcal{X}, d)$, we define the *Voronoi Region* $R(x)$ of a point $x \in S$ to be

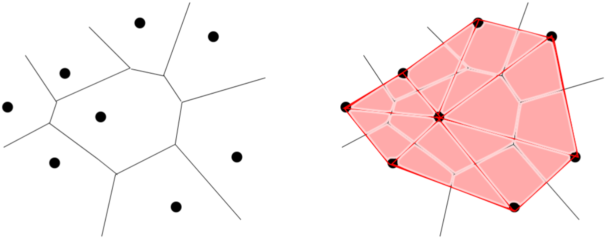
$$R(x) = \{y \in \mathcal{X} | d(x, y) \leq d(s, y), \forall s \in S \backslash \{x\}\}$$

and the *Voronoi Diagram of* S, $V(S)$, to be the set

$$V(S) = \{R(x) | x \in S\}$$

The Voronoi Diagram forms a closed cover of S and so we can consider the nerve of $V(S)$, which is called the *Delaunay Complex*.

Delaunay Complex: Example

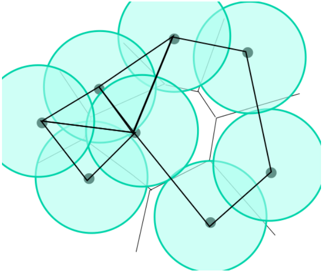


(a) Voronoi Region (b) Delaunay Complex

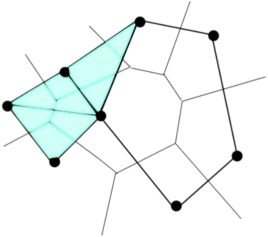
This is a very “filled” simplicial complex. Is there a way to capture more of the intricacies of the data?

Alpha Complex: Example

The Alpha complex is obtained by taking ϵ-balls around each point in S, intersecting these balls with the Voronoi regions, and taking the nerve of this cover.



(c) Alpha Balls and Nerve



(d) Alpha Complex

Notice that the Alpha complex is a subcomplex of the Delaunay complex. If $\epsilon = \infty$, then the alpha complex *is* the Delaunay complex.

*n*th Chain Group

We now turn to simplicial homology to analyze the shape of a given simplicial complex. Here, we do not want to differentiate between different orientations of simplicies and so we will denote an *oriented simplex* as $[\sigma]$ for simplex σ.

Definition

Given a simplicial complex $\mathcal{K}$, the *nth Chain Group* $C_n(\mathcal{K})$ is the free Abelian group on $\mathcal{K}$'s oriented n-simplicies.

Thus, elements in $C_n(\mathcal{K})$ can be expressed as finite sums...

$$c = \sum_i c_i[\sigma_i]$$

where $\sigma_i \in \mathcal{K}$ are n-simplicies and $c_i \in \mathbb{Z}$ for all i.

Boundary Homomorphism

Given the Chain Groups, we can define the n-boundary homomorphism $\partial_n : C_n \to C_{n-1}$ to be,

$$\partial_n[v_0, ..., v_n] = \sum_{i=0}^{n} (-1)^i [v_0, ..., \hat{v}_i, ..., v_n]$$

where $\hat{v}_i$ denotes removing v_i from the vertex sequence.

Example: Triangle (on board)

$$\partial_1([a, c] - [b, c] + [b, a]) = ([c] - [a]) - ([c] - [b]) + ([a] - [b]) = 0$$

Thus, $[a, c] - [b, c] + [b, a] \in \ker \partial_1$.

Simplicial Homology

Note that $\partial_n \circ \partial_{n+1} \equiv 0$ (exercise) and therefore, the boundary operator connects the chain groups into a chain complex:

$$\ldots \to C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \to \ldots$$

on which we can analyze the *nth homology group* H_n,

$$H_n = \ker \partial_n / \mathrm{im} \partial_{n+1}$$

Elements in $\ker \partial_n$ are *cycles* and elements in $\mathrm{im}\partial_{n+1}$ are *boundaries*.

Betti Numbers

We can consider the *Betti Number* of the *n*th homology group,

$$\beta_n = \text{rank}(H_n) = \min\{|X| : X \subseteq H_n, \langle X \rangle H_n\}$$

This gives the number of *n*-dimensional holes in our space. Note that β_n will be non-zero for only finitely many n. Additionally, since $\mathcal{K}$ is finite, H_n is finitely generated and therefore $\beta_n < \infty$ for all n.

A nice feature due to the Euler-Poincaré formula states that

$$\chi(\mathbb{X}) = \sum_{n=1}^{\infty} (-1)^n \beta_n$$

where χ is the Euler characteristic.

Single-Scale to Multi-Scale

Recall that we derived the alpha complex from a single parameter ϵ. That is, we constructed balls of radius ϵ around our data points and intersected these balls with the Voronoi regions at each point. Taking the nerve of this cover yielded the alpha complex.

Therefore, the invariants we calculated were dependent on our choice of ϵ. *Persistence Homology* allows one to vary the value of ϵ and look for simplicies and invariants which do not change with varying ϵ. That is, features which persist through changes in scale.

Discussion of persistence homology will be left for another talk.

Thanks!

References

Afra Zomorodian (2011)
Topological Data Analysis
Proceedings of Symposia in Applied Mathematics